基于改进粒子群算法的云计算产业联盟知识搜索算法研究*

高长元 [1,2] 于建萍 [1] 何晓燕 [1,2]
[1](哈尔滨理工大学管理学院 哈尔滨 150040)
[2](哈尔滨理工大学高新技术产业发展研究中心 哈尔滨 150040)

摘要: 【目的】利用改进的粒子群算法进行云计算产业联盟知识搜索，提高搜索的准确率和效率。【方法】首先利用MapReduce中Map函数对粒子分组实现并行化处理，再运用Reduce函数对粒子搜索的结果进行归约，缩短搜索的时间。在粒子搜索过程中，根据小组内最优位置的平均值进行小组内粒子的信息交互，避免算法早熟收敛于一个局部最优值。【结果】通过三组仿真实验对改进的粒子群算法和标准粒子群算法进行对比分析，结果表明改进的粒子群算法在效率与准确率方面均具有明显的优越性。【局限】样本数据存在干扰数据，有待改进。【结论】该方法能提高云计算产业联盟知识搜索的准确性，并提升搜索效率。

关键词: 云计算产业联盟 知识搜索 粒子群优化算法 MapReduce
分类号: G250.2 F272.4

1 引 言

云计算自提出以来发展迅猛，世界各国对它广阔的市场前景和巨大的产业机遇给予高度关注。根据赛迪智库发布的《云计算产业发展白皮书(2015 版)》显示，2015 年全球云计算服务市场规模达到 1 800 亿美元，增长 18%，全球 IT 公司纷纷向云计算转型，云计算产业规模高速增长。云计算产业联盟的成立正是为了致力于云计算产业服务，推动云计算生态系统和云计算产业发展。云计算产业联盟是聚集云计算产业链上下游重点企业，以合同或契约关系结成的联合体，联盟成员之间互相信任、共享知识，提高联盟内知识共享利用率，实现云计算产业联盟和成员利益最大化的新型产业组织。翟丽丽等指出组建云计算产业联盟的一个重要的原因是联盟成员相互信任，能够借助联盟的力量实现知识共享，从而弥补成员自身的知识缺口，优化联盟成员的知识配置[1]。云计算产业联盟成员依靠云计算技术各自形成存放知识以及外部知识的可共享云数据库，通过云计算产业联盟知识平台对云数据库的访问，实现联盟内的知识搜索。因此，联盟成员如何在云数据库中迅速而准确地查到所需要的知识，是亟需解决的问题。

为解决上述问题，并有效针对云计算产业联盟中知识的海量性和多样性及云数据库的分布式和动态可扩展性等特点[2]，需要运用一种搜索算法使联盟成员能够准确而又快速地获取知识。算法的选择主要遵从以下两个方面:

(1) 搜索时间问题。随着云计算产业联盟的运营及成员的增加，云数据库存量和数量不断变大变多，为联盟成员搜索造成困难，所以用降低最优值精度的方法来提高计算效率[3]。

(2) 全局最优性。云计算产业联盟云数据库是庞大的资源池，普通的搜索算法，如蚁群算法，因收敛于局部最优解[4]，不能为联盟成员搜索到更精准的知识。

粒子群是一种启发式算法，它能同时解决静态及动态的组合优化问题[5]。云计算产业联盟成员知识搜

通讯作者: 高长元，ORCID: 0000-0003-1006-8343，E-mail: gaocy2002@126.com。
*本文系黑龙江省自然科学基金项目“黑龙江省移动云计算联盟商业模式研究”(项目编号: G201301)的研究成果之一。

索过程与粒子群中鸟找食行为的原理十分切合。然而云计算产业联盟中云数据库众多且分布复杂，仅用标准粒子群算法进行知识搜索，粒子之间缺乏信息沟通和合作，最后陷入局部最优解，不能全面访问云数据库。在避免粒子群算法早熟方面，前人提出许多改进方法，对粒子群优化算法的参数进行自适应选择策略，如采用动态非线性函数控制参数[6]、利用 Q-学习调节参数[7]、采用自适应评论策略调节参数[8]等。此外，王燕燕等提出一种动态分组的粒子群优化算法，划分的子种群数目从特定集合中随机选取，虽然能够提高整体求解质量，但对于复杂问题仍存在不足[9]。众多专业学者越来越关注粒子群算法，由于求解优秀，粒子群算法已运用于云计算领域。李志洁等利用改进的粒子群，处理网格资源分配问题，从而实现存储约束条件下网格资源优化分配[10]。李媛媛等提出一种快速收敛的改进粒子群优化算法来进行云平台的调度，能够以较高的效率和准确率完成所有任务[11]。上述表明，改进的粒子群算法在解决云计算领域的问题上存在优势，并受到一致认可，但是研究多是对粒子群算法本身的参数进行修改，没有考虑云计算产业联盟和联盟内知识的分散、多样等特性。

基于以上分析，本文针对云计算产业联盟内知识及云数据库的特点，提出一种基于 MapReduce 的改进粒子群算法。一方面针对云计算产业联盟中知识的动态增量的特点，结合 MapReduce 函数完成对粒子群的分组，实现并行化处理，提高搜索效率，避免使用随机生成所造成初始位置具有片面性的问题；另一方面针对粒子群原有的算法进行改进，对粒子分组利用小组内最优位置的平均值进行小组内粒子的信息交互，避免算法收敛于一个局部最优值，从而提高搜索效率与准确率。

2 云计算产业联盟知识搜索分析

2.1 云计算产业联盟的成员分析

联盟的成员包括：软件提供商、网络运营商、硬件设施提供商、内容提供商、科研机构和云计算服务提供商。云计算产业联盟成员通过网络及应用软件，向云计算服务提供商发出请求，并将所需服务返回给成员。软件服务提供商负责为其他各成员提供软件支持，并为整个联盟运行提供软件应用，使成员通过软件应用连接云服务端。网络运营商主要为其他各成员提供网络服务、内容服务、增值服务，并为整个联盟的运行提供网络支持，使用户依托网络接入云。硬件设施提供商主要是为云计算服务提供商提供硬件设施，帮助搭建云平台、云应用等，为整个联盟的运行提供硬件基础。科研机构主要提供理论与技术支持，包括联盟管理、联盟运营以及风险控制与利益分配等设计。云计算服务提供商是整个联盟运行的核心，提供云资源服务、云平台服务以及云应用服务，它需要依托各成员的支持，并将云服务提供给各成员。

根据云计算产业联盟的运行情况以及联盟成员自有的属性，云计算产业联盟知识的类型主要分为以下类别：运营商知识、硬件设备知识、软件知识、内容知识、云计算服务知识、科研知识和联盟创新知识，如表 1 所示。

表 1 云计算产业联盟知识类型

知识类型	描述
运营商知识	移动、联通、电信在运行中所产生的知识
硬件设备知识	制造硬件过程中产生的知识
软件知识	软件设计、运行、维护的知识
内容知识	文字、图像、音频和视频等各种媒体内容
云计算服务知识	包括云平台知识、云资源知识以及云应用知识
科研知识	研究院、高校研究提供理论与技术支持
联盟创新知识	联盟在运营中所产生的知识

2.2 云计算产业联盟知识特性分析

(1) 分散性：云计算产业联盟成员广泛分布在不同地理位置，成员的知识均保存在各自的云数据库中，根据知识类型不同的组织形态存储知识。

(2) 多样性：云计算产业联盟中的知识，不仅包括运营商知识、硬件设备知识、软件知识、内容知识、云计算服务知识以及科研知识，还包括云计算产业联盟创新的知识、维护联盟的知识、管理技能知识以及联盟的公共关系知识。

(3) 海量性：云计算产业联盟中的各种知识无时无刻均被存储在各个不同的云数据库，此外联盟成员的知识进行各种交互与协作都会产生海量的交互、交易知识。

(4) 增量性：随着云计算产业联盟的发展和运作，联盟中的知识以动态增长的形式大量产生和积累，联盟成员通过进行知识交互，成员本身的知识存量也在

不断增长。

(5) 资产性: 知识可以看作一种无形的资产, 不同企业通过云计算产业联盟平台, 提交各自的知识, 共同服务于联盟, 不断创新, 将知识作为一种服务或者是商品提供给不同的用户。所以知识的价值在于不断开发、挖掘、处理等过程中创新出的知识商品(服务)。

云计算产业联盟成立的主要目标是实现知识共享, 但是云计算产业联盟中知识类型繁多, 并且联盟知识的多样性、海量性和增量性等特点, 决定了联盟成员进行知识共享的困难。联盟成员在联盟中获取知识, 提高成员自身的竞争力和创新力, 从而在激烈的市场竞争中具备优势, 所以联盟内知识的搜索就显得尤为重要。

2.3 云计算产业联盟知识搜索优化过程分析

云计算产业联盟内的成员拥有形成云环境的技术基础, 联盟内的成员企业提供自己的基础设施或服务器, 组成一个用于存放各成员知识以及外部知识的云数据库, 并共同搭建云计算产业联盟知识平台。通过知识共享平台实现对云数据库的搜索, 使联盟内成员能够快速便捷地获取所需知识。联盟成员可免费提供给其他成员的知识以及需要付费才能获取的知识通过本体化表达, 利用云存储技术存储在云数据库中, 不同类型的知识用不同的本体表达。联盟成员进行知识搜索时, 根据自己的搜索需求在引擎数据库中选取相应的知识类型的本体表现形式作为粒子初始所携带的知识, 通过 API 接口, 可以实现对云数据库的访问。

云计算产业联盟的成员众多, 导致存放成员知识的云数据库数量巨大, 并且由于成员的加入和退出, 使得联盟的云数据库数量是动态变化的。而联盟云数据库中存放的知识具有多样性、海量性和增量性等特性。由于标准粒子群算法随机生成初始搜索位置, 导致标准粒子群算法在联盟中按照原来的串行搜索方式进行知识搜索时计算难度大, 搜索时间长。因此引入处理海量知识有效的 MapReduce 函数, 准确计算出粒子的初始化位置, 帮助其实现知识搜索的并行化处理。此外, 标准粒子群算法进行搜索时要所有的粒子同时更新, 且需要大量的最优位置共享和交互才能完成迭代, 但是在云计算产业联盟分布式环境下搜索就显得尤为困难。所以在云计算产业联盟环境下, 提出对粒子群进行分组, 根据小组内最优位置的平均值进行小组内粒子的信息交互。本文基于云计算产业联盟背景, 在此基础上采用 MapReduce 函数解决其并行化处理及动态性增强的问题, 同时改进标准粒子群算法。综合考虑粒子群与 MapReduce 函数, 创新性地提出基于改进粒子群算法的云计算产业联盟知识搜索过程, 如下:

(1) 根据云计算产业联盟中实时云数据库的个数, 利用 MapReduce 中 Map 函数对整个粒子群进行分组, 由原始的串行搜索改进为并行搜索, 减少搜索的时间, 提升搜索的效率。

(2) 粒子进入云数据库后, 用携带的知识与云数据库中的知识不断进行对比, 粒子根据联盟成员的需求选择最适合的知识, 同时以小组内最优位置的平均值作为粒子的移动依据。

(3) 当粒子搜索结束后, MapReduce 中 Reduce 函数, 对不同组的粒子搜索结果进行归约, 并将归约后的一个或者多个结果返还给联盟成员, 具体搜索流程如图 1 所示。

3 基于粒子群算法云计算产业联盟知识搜索建模

3.1 MapReduce 函数构建

利用粒子群算法进行搜索, 粒子的个数越多, 算法的效率越高, 效果也更为显著。但是, 粒子的个数越多, CPU 被占用得越多。所以利用云计算 MapReduce 分布式计算和分集存储的优势, 能够给粒子足够的内存空间, 同时加快迭代次数。首先运用 Map 函数对整个粒子群进行分组, 并分配到不同的节点上执行, 实现并行化处理, Reduce 函数实现搜索结果的归约, 这里的节点即为云数据库中一个独立的数据库。其解决了传统粒子群算法中由于单向局部收敛速度过快而导致没有收敛到全局最优的问题, 并且使搜索时间得到有效提升, 另一方面也更加适合处于动态变化中的联盟知识。本文首先利用 Map 函数根据节点的个数将一个完整的粒子种群划分成子群, 并分配到不同的云数据库, 实现并行化处理。粒子搜索到结果后, 再利用 Reduce 函数将不同小组的粒子进行合并, 具体过程如图 2 所示。

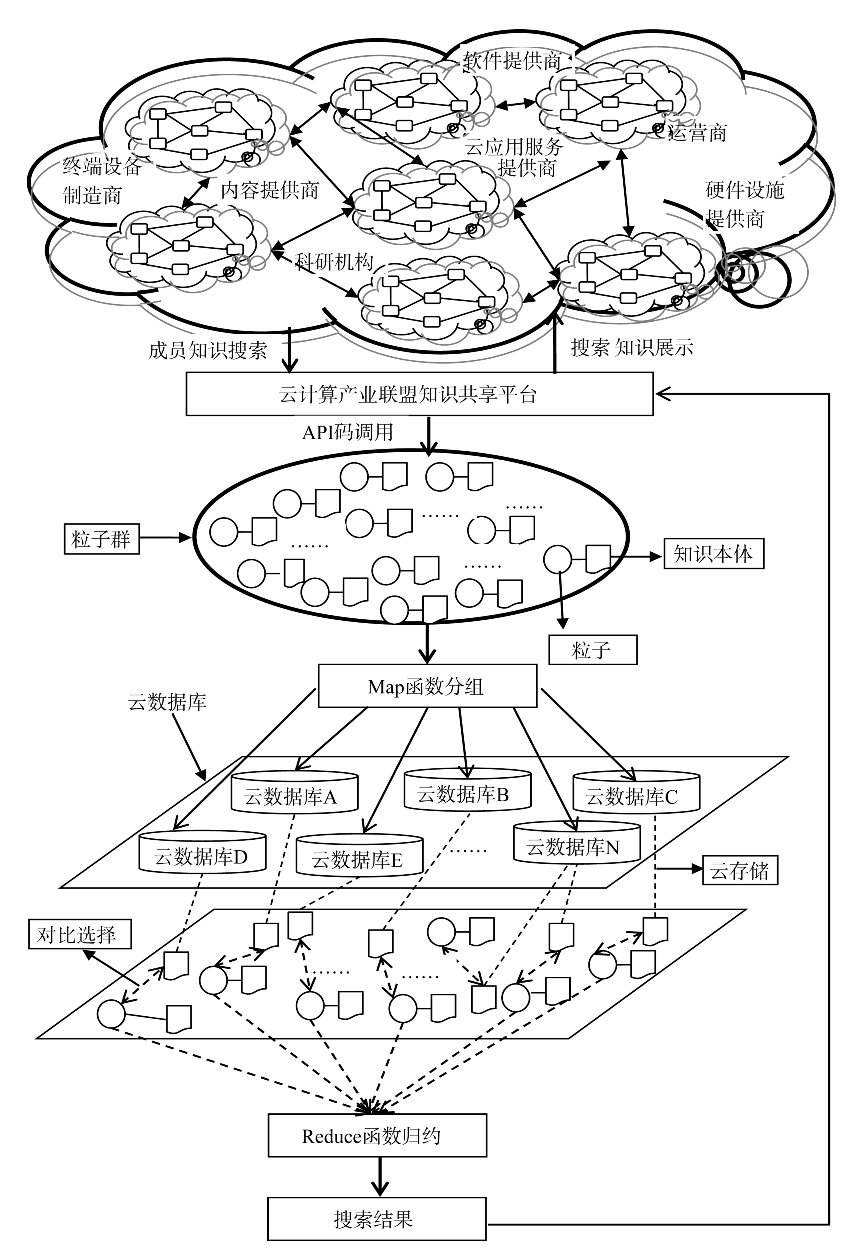


图 1　云计算产业联盟知识搜索优化过程图

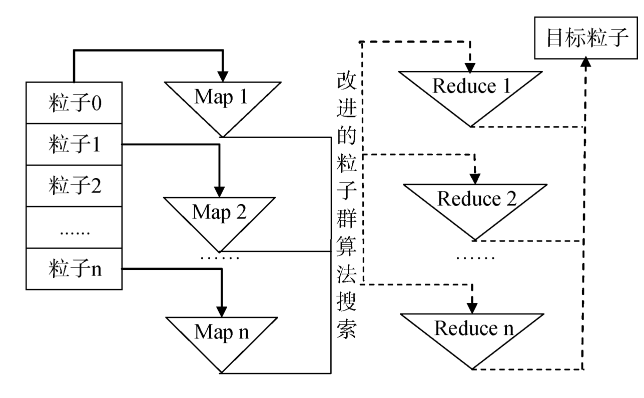


图 2　MapReduce 并行化处理过程

3.2　基于 MapReduce 的粒子群算法改进

粒子群优化算法(Particle Swarm Optimization, PSO)是由J.Kennedy博士和 R.C.Eberhart 博士于1995年共同提出的，因其算法简单、需要调节的参数较少、较强的全局优化能力和高效等特点，立即成为研究热点[12]。标准粒子群算法的数学描述如下[13]:

m 个粒子在 D 维的空间中进行搜索，第 i 个粒子的位置用向量 $x_i=[x_{i1},x_{i2},\cdots,x_{iD}]$ 表示，飞行速度用 $v_i=[v_{i1},v_{i2},\cdots,v_{iD}]$ 表示，第 i 个粒子搜索到的最优位置为 $pbest_i=[p_{i1},p_{i2},\cdots,p_{iD}]$，整个群体搜索到的最

优位置为 $gbest_i = [pg_{i1}, pg_{i2}, \cdots, pg_{iD}]$，则每一个粒子的速度和位置更新如公式(1)和公式(2)所示。

$$v_i^{k+1} = \omega v_i^k + c_1 r_1 (pbest_i - x_i^k) + c_2 r_2 (gbest_i - x_i^k) \quad (1)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \quad (2)$$

其中，$i = 1, 2, \cdots m$，分别代表不同的粒子。学习因子 c_1 和 c_2 是两个非负数，r_1 和 r_2 是介于 $[0,1]$ 的一个随机数，k 为迭代次数，ω 是惯性权重[14]。

标准的粒子群算法进行搜索时每一次迭代都需要所有的粒子同时更新才能完成，但是在分布式特征最为明显的云计算环境中搜索就显得尤为困难。而且在粒子进行迭代时，每一次移动都需要所有粒子共享最优位置，并且随时都可能更新最优值，即粒子需要大量的最优位置共享和交互才能迭代，这并不适合云计算环境。因此在云计算分布式环境下，粒子如果能够以异步进行迭代，降低粒子间共享位置的数据传输，是解决问题的关键。标准的粒子群算法虽然具有鲁棒性强等特点，但是算法在搜索初期，粒子的分布较为分散，尽管能够在较大的空间中进行搜索，但需要的搜索时间长。当粒子搜索一段时间后，由于粒子的历史最优位置和种群的最优位置接近一致，造成粒子行动迟缓，丧失寻优能力，算法早熟并收敛于局部最优值。所以，算法改进的另一个方面是要保持粒子种群的多样性[15]，提高粒子的搜索能力。

本文对算法的改进如下：粒子群分组后，每组粒子分别独立进行搜索，对于第 j 组粒子，其速度和位置更新如公式(3)和公式(4)所示。

$$v_i^{k_j+1} = \omega v_i^{k_j} + c_1 r_1 (pbest_i^{at} - x_i^{k_j}) + c_2 r_2 (gbest_i^{k_j} - x_i^{k_j}) \quad (3)$$

$$x_i^{k_j+1} = x_i^{k_j} + v_i^{k_j+1} \quad (4)$$

其中，k_j 表示第 j 组粒子的迭代次数，$gbest_i^{k_j}$ 表示第 j 组粒子的全局极值。$pbest_i^{at}$ 表示在 k 时刻小组内最优位置的平均值，其余参数含义同标准粒子群算法。

本文对算法的改进一方面是为了让算法在云计算分布环境下得以实现，另一方面是利用云计算优势提高算法速度，使早熟收敛的问题及时得到解决，并且难以从局部最优中跳出来的问题也得以改善。

3.3　改进的粒子群算法步骤

在改进的粒子群算法中，粒子作为携带知识的一种载体，通过与数据库中不同类型的知识进行对比选择实现搜索。当算法结束时，得到的最优粒子即为云计算产业联盟成员所需的知识，搜索完成。

改进的 PSO 算法(CPSO)知识搜索过程如下。

(1) 根据联盟成员的搜索条件，从引擎数据库中确定和联盟成员输入关键字相关的本体化表达的知识大类作为初始化种群。

(2) 利用 MapReduce 中的 Map 函数根据云数据库的个数对粒子群进行映射分组。

(3) 更新粒子的速度和位置。按照公式(3)和公式(4)改变每组的粒子速度和位置。

(4) 知识选择。在每个小组中，比较各组粒子适用值及其小组内最优位置的平均值 $pbest_i^{at}$，若粒子当前适用值优于 $pbest_i^{at}$，则重新计算当前粒子组的位置并重置 $pbest_i^{at}$ 的值；比较粒子的适用值及小组内最优值 $gbest_i^{k_j}$，若粒子适用值优于 $gbest_i^{k_j}$，则用当前粒子位置重置 $gbest_i^{k_j}$ 的值。

(5) 当粒子搜索到结果或算法的迭代次数达到最大迭代次数 T-max 时，利用 Reduce 函数对不同种群粒子最后产生的最优解进行归约。

(6) 将最优解或者最优解集返回给联盟成员。

4　仿真实验

为更好验证改进后的粒子群算法在云计算产业联盟知识搜索过程中的有效性，本文进行三组仿真实验。分别对改进的粒子群算法、改进的粒子群算法在云数据库搜索中的平均吞吐量及分组延迟中的比较和改进的粒子群算法在云数据库中的搜索进行仿真分析。第一个实验仿真软件为 MATLAB 7.0；第二个实验选择南加州大学情报局(South California University ISI)研究院的 NS-2 作为模拟平台[16]；而第三个实验选择 Linux 系统 Hadoop 2.0 搭建云数据库和云平台。实验仿真数据来自加州大学开发 UCI KDD Archive，用于机器学习和知识发现研究的数据库平台。根据云计算产业联盟及联盟内知识的特性，选择符合联盟知识特性的数据集，进行仿真模拟。

4.1　改进的粒子群仿真分析

用三种 Benchmark 函数分别对标准粒子群算法和改进的粒子群算法进行测试。

(1) Rastrigin 函数

$$f_1(x)=\sum_{i=1}^{n}(x_i^2-10\cos(2\pi x_i)+10), x_i\in[-5.12,5.12]$$

(2) Griewank 函数

$$f_2(x)=\frac{1}{4000}\sum_{i=1}^{n}x_i^2-\prod_{i=1}^{n}\cos\left(\frac{xi}{\sqrt{i}}\right)+1, x_i\in[-600,600]$$

(3) Rosenbrock 函数

$$f_3(x)=\sum_{i=1}^{n}[100(x_{i+1}-x_i^2)^2+(1-x_i)^2], x_i\in[-30,30]$$

Rastrigin [9]和 Griewank [14]都是多峰函数，利用一般的算法搜索最优值有很大的难度；Rosenbrock 是单峰函数，函数取值走势平缓，难以搜索到最优解。三个函数的维数都为 10，最优解都为 0。

函数测试迭代次数为 1 000，三个函数都独立运行 100 次。模拟 10 个计算单元，每个计算单元有 50 个粒子，所以种群规模等效于 500 个粒子。同样标准粒子群中粒子个数也为 500。标准 PSO 算法的参数设置为：$c_1=2.05$；$c_2=2.00$；$\omega=0.77$。为了方便比较，优化算法的 c_1、c_2、ω 参数取值与标准算法相同。实验结果如表 2 所示。

表 2　函数实验结果比较

函数	算法来源	平均值	最好值	最差值
Rastrigin	PSO	18.34	6.58	29.95
0	CPSO	5.20	1.89	8.98
Griewank	PSO	0.81	0.36	1.15
0	CPSO	0.23	0.05	0.45
Rosenbrock	PSO	200.63	13.95	908.72
0	CPSO	5.89	0.18	7.19

从上述的实验结果可以看出，在相同粒子的情况下进行寻优时，利用改进的粒子群算法对多峰函数 Rastrigin 寻优，其平均值、最好值和最差值均有提升，平均值从 18.34 提升到 5.20，全局搜索能力有明显的提高；多峰函数 Griewank 的平均值、最好值和最差值也有提升，但是由于函数本身值域较为集中，平均值仅从 0.81 提升到 0.23；Rosenbrock 函数是复杂的非凹、病态单峰函数，所以全局搜索能力极为重要，函数原本提供的信息少，传统的粒子群算法很难在短时间内辨别搜索的方向，从而陷入局部最优，也因此造成单次结果差异很大。改进的粒子群算法因全局遍历性和分布均匀，能够在短时间内找到正确的搜索方向，进而提升算法跳出局部最优解的能力。

4.2　云数据库吞吐量及分组延迟仿真分析

选择实验网络数据恰当与否取决于许多以非线性和不可预测的方式相互作用的影响程度，所以选择一个有意义的测试环境是非常困难的：遵循的方法是定义可调成分的各种分类组成的有限集合[17]。选择南加州大学情报局研究院的 NS-2 作为该实验平台。选用吞吐量及数据分组延迟作为对算法有效性的衡量标准，实验数据是 50 次实验的平均值，每次仿真时间为 1 000 个虚拟秒。

从图 3 的实验结果来看，改进的粒子群算法更优，每秒正确传输的比特数保持在 1000Kb/s，而标准的粒子群算法明显低于改进的粒子群算法。从图 4 的实验结果来看，改进的粒子群算法时间延迟较低。

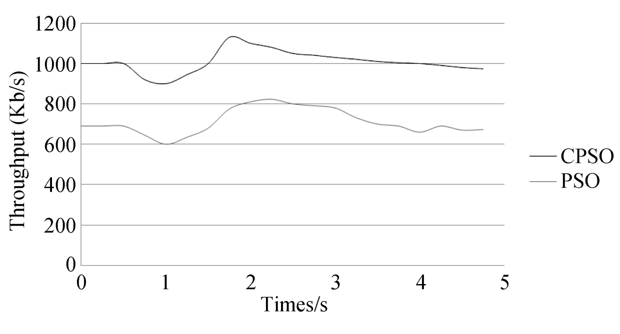


图 3　平均吞吐量

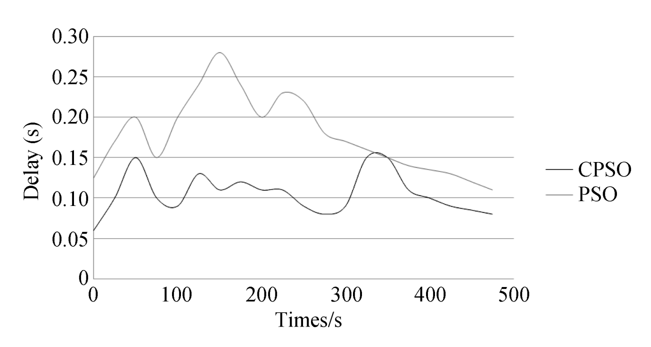


图 4　分组延迟的经验分布

4.3　搜索时间和搜索准确率分析

本文研究的是云计算产业联盟知识搜索，为验证知识搜索的准确率和时效性，搭建云计算产业联盟知识搜索实验环境，建立虚拟机。在虚拟机下，搭建 20 个 Linux 环境，每个 Linux 系统下采用 Hadoop 2.0 搭建用来表示云计算产业联盟成员存储知识的云数据库和进行知识搜索的云平台。

在第一组仿真实验中，将标准粒子群算法(PSO)

和改进的粒子群算法(CPSO)依次在搭建好的云计算产业联盟环境中进行知识搜索。通过图 5 可以看出，在进行知识搜索时，改进的粒子群算法的准确率更高，耗时更小，这是因为在模拟的云计算产业联盟环境中云数据库众多，标准的粒子群初始搜索位置随机生成，搜索方式是串行搜索，需要耗费更多时间。在搜索的后期，粒子的历史最优位置和群体的最优位置趋近一致，使得粒子移动速度过慢，基本丧失寻优能力，准确率不高。在第二组仿真实验中，将文献[9]提出的动态分组的粒子群算法(DGPSO)和改进的粒子群算法(CPSO)依次在搭建好的云计算产业联盟环境中进行知识搜索，通过图 6 可以看出改进的粒子群算法的准确率仍然更高，且耗时更小。通过图 5 和图 6 的对比可知，动态分组粒子群算法的准确率要比标准粒子群算法高，性能有所提升，耗时也更低，但是在云计算产业联盟复杂的环境下，搜索能力还是没有本文提出的改进的粒子群算法强。

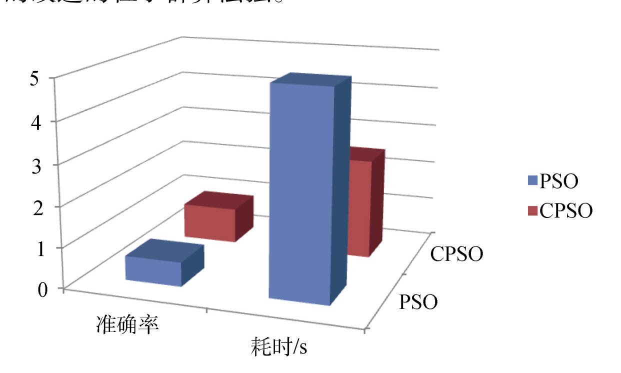


图 5　第一组仿真实验结果

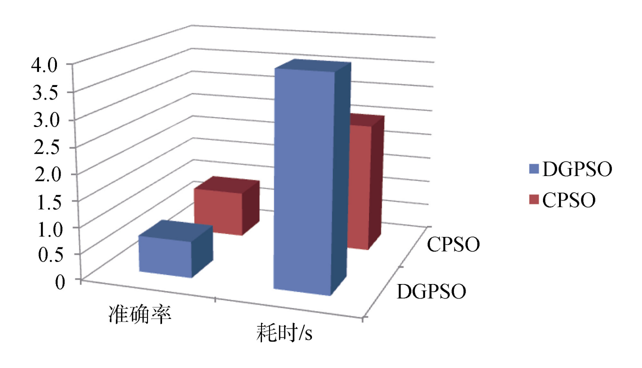


图 6　第二组仿真实验结果

5　结　语

为解决由于云计算产业联盟知识特性导致联盟成员难以在云数据库进行知识搜索的问题，本文提出一种改进的粒子群算法对联盟云数据库实现知识搜索。在粒子群算法的基础上结合 MapReduce 并行计算的思想，减少原本算法在对云计算产业联盟大量丰富且动态的知识进行搜索的时间；同时将粒子分组，更好地提高云计算产业联盟知识搜索的准确性。最后通过三组仿真实验对改进的粒子群算法和标准粒子群算法进行对比分析，结果表明改进的粒子群算法在效率与准确率方面均具有明显的优越性。在下一步的研究中，将致力于云计算产业联盟知识搜索结果的筛选匹配方面，从而保障云计算产业联盟成员进行有效知识获取。

参考文献：

[1] 翟丽丽，柳玉凤，王京，等. 软件产业虚拟集群企业间信任进化博弈研究[J]. 中国管理科学，2014，22(12): 118-125. (Zhai Lili, Liu Yufeng, Wang Jing, et al. Research on Evolutionary Game on Trust Among Software Industrial Virtual Cluster's Enterprises [J]. Chinese Journal of Management Science, 2014, 22(12): 118-125.)

[2] 史恒亮，任崇广，白光一，等. 自适应蚁群优化的云数据库动态路径查询[J]. 计算机工程与应用，2010，46(9): 10-12. (Shi Hengliang, Ren Chongguang, Bai Guangyi, et al. Cloud Database Dynamic Route Query Based on Self-adaptive Ant Colony Optimization [J]. Computer Engineering and Applications, 2010, 46(9): 10-12.)

[3] Hiremath N C, Sahu S, Tiwari M K. Multi Objective Outbound Logistics Network Design for a Manufacturing Supply Chain[J]. Journal of Intelligent Manufacturing, 2013, 24(6): 1071-1084.

[4] 苏屹，李柏洲，刘晓静. 基于蚁群算法的 R&D 支出拟合模型研究[J]. 情报杂志，2012，31(5): 198-201. (Su Yi, Li Baizhou, Liu Xiaojing. A Model Fitting Analysis of the R&D Expenditure Based on Ant Colony Algorithm[J]. Journal of Intelligence, 2012, 31(5): 198-201.)

[5] 王洪峰，王娜，汪定伟，等. 一种求解多峰优化问题的改进 Species 粒子群算法[J].系统工程学报，2012，27(6): 854-864. (Wang Hongfeng, Wang Na, Wang Dingwei, et al. Improved Species-based Particle Swarm Optimizer for Multi-modal Optimization Problems [J]. Journal of Systems Engineering, 2012, 27(6): 854-864.)

[6] Chatterjee A, Siarry P. Nonlinear Inertia Weight Variation for Dynamic Adaptation in Particle Swarm Optimization[J]. Computers & Operations Research, 2006, 33(3): 859-871.

[7] Stefan J, Martin M. A Hierarchical Particle Swarm Optimizer and Its Adaptive Variant[J]. IEEE Transactions on Systems,

Man and Cybernetics-Part B: Cybernetics, 2005, 35(6): 1272-1282.
[8] Doctor S, Venayagamoorthy G K. Improving the Performance of Particle Swarm Optimization Using Adaptive Critics Designs[J]. Swarm Intelligence Symposium, 2005, 9(9): 393-396.
[9] 王燕燕, 葛洪伟, 王娟娟, 等. 一种动态分组的粒子群优化算法[J]. 计算机工程, 2015, 41(1): 180-185, 189. (Wang Yanyan, Ge Hongwei, Wang Juanjuan, et al. A Particle Swarm Optimization Algorithm of Dynamic Grouping [J]. Computer Engineering, 2015, 41(1): 180-185, 189.)
[10] 李志洁, 刘向东, 段晓东. 改进粒子群算法在网格资源分配中的优化[J]. 计算机集成制造系统, 2009, 15(12): 2375-2382. (Li Zhijie, Liu Xiangdong, Duan Xiaodong. Optimization of Grid Resource Allocation Using Improved Particle Swarm Optimization Algorithm [J]. Computer Integrated Manufacturing Systems, 2009, 15(12): 2375-2382.)
[11] 李媛媛, 曲雯毓, 栗志扬, 等. 一种快速收敛粒子群优化算法在云计算中应用[J]. 华中科技大学学报: 自然科学版, 2012, 40(S1): 34-37. (Li Yuanyuan, Qu Wenyu, Li Zhiyang, et al. Particle Swarm Optimization Algorithm with Fast Convergence Applied in Cloud Computing[J]. Journal of Huazhong University of Science and Technology: Natural Science Edition, 2012, 40(S1): 34-37.)
[12] Shelokar P S, Siarry P, Jayaraman V K. Particle Swarm and Ant Colony Algorithms Hybridized for Improved Continuous Optimization[J]. Applied Mathematics and Computation, 2007, 188(1): 129-142.
[13] 范成礼, 邢清华, 付强, 等. 求解非线性双层规划问题的混合变邻域粒子群算法[J]. 系统工程理论与实践, 2015, 35 (2): 473-480. (Fan Chengli, Xing Qinghua, Fu Qiang, et al. A Hybrid Intelligent Algorithm by Combining Particle Swarm Optimization with Variable Neighborhood Search for Solving Nonlinear Bilevel Programming Problems [J]. Systems Engineering-Theory & Practice, 2015, 35(2): 473-480.)
[14] 罗宏. 基于改进粒子群算法的图书封面搜索引擎研究[J]. 情报科学, 2014, 32 (4): 106-108, 121. (Luo Hong. Research on the Book Covers Search Engine Based on Improved PSO[J]. Information Science, 2014, 32 (4): 106-108, 121.)
[15] 李国辉, 冯明月, 易先清. 基于分群粒子群优化的传感器调度方法[J]. 系统工程与电子技术, 2010, 32(3): 598-602. (Li Guohui, Feng Mingyue, Yi Xianqing. Sensor Scheduling Method Based on Grouping Particle Swarm Optimization[J]. Systems Engineering and Electronics, 2010, 32(3): 598-602.)
[16] 叶晓国. 基于 NS-2 的无线传感器网络仿真模块扩展方法的研究[J]. 计算机研究与发展, 2011, 48 (S2): 302-306. (Ye Xiaoguo. NS-2- Based Simulation Module Extension Method for Wireless Sensor Networks[J]. Journal of Computer Research and Development, 2011, 48 (S2): 302-306.)
[17] Liu Y, Chen J, Wang J. On Counting 3-D Matching of Size K[J]. Algorithmic, 2009, 54(10): 530-543.

作者贡献声明:

高长元: 提出研究思路, 设计研究方案, 论文最终版本修订;
于建萍: 方案设计和修订, 论文起草以及修订;
何晓燕: 方案设计, 论文多版本修订。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据见期刊网络版 http://www.infotech.ac.cn。
[1] 高长元, 于建萍, 何晓燕. text.xlsx. 测试数据.
[2] 高长元, 于建萍, 何晓燕. train.xlsx. 训练数据.

收稿日期: 2016-06-27
收修改稿日期: 2016-11-27

Knowledge Search for Cloud Computing Industry Alliance: An Algorithm Based on Improved Particle Swarm Optimization

Gao Changyuan[1, 2]　Yu Jianping[1]　He Xiaoyan[1, 2]
[1](College of Management, Harbin University of Science and Technology, Harbin 150040, China)
[2](High-tech Industrial Development Research Center, Harbin University of Science and Technology, Harbin 150040, China)

Abstract: [**Objective**] This paper uses an algorithm based on the improved particle swarm optimization to conduct knowledge search for cloud computing industry alliance, aiming to improve its accuracy and efficiency. [**Methods**] First, we utilized the Map function of the MapReduce model to process particle grouping. Secondly, we used the Reduce function to shorten the particle search result lists and search time. Lastly, the information interaction of the particles was decided by the average value of the optimal position within each group, which avoided the premature convergence of using a local optimal value. [**Results**] We compared the performance of the improved algorithm with the standard ones by three rounds of simulation experiments. We found that the improved particle swarm algorithm was superior in efficiency and accuracy. [**Limitations**] There is some noisy data in the sample. [**Conclusions**] The proposed algorithm could improve the accuracy and efficiency of knowledge search for the cloud computing industry alliance.

Keywords: Cloud Computing Industry Alliance　Knowledge Search　Particle Swarm Optimization Algorithm　MapReduce

---